



If You Build It, Will They Fund?

Making Research Data Management Sustainable

Ricky Erway and Amanda Rinehart

If You Build It, Will They Fund? Making Research Data Management Sustainable

Ricky Erway

OCLC Research

Amanda Rinehart

The Ohio State University Libraries



© 2016 OCLC Online Computer Library Center, Inc.

This work is licensed under a Creative Commons Attribution 4.0 International License.

<http://creativecommons.org/licenses/by/4.0/>



January 2016

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 1-55653-501-5 (978-1-55653-501-7)

OCLC Control Number: 930726586

Please direct correspondence to:

Ricky Erway

Senior Program Officer

erwayr@oclc.org

Suggested citation:

Erway, Ricky, and Amanda Rinehart. 2016. *If You Build It, Will They Fund? Making Research Data Management Sustainable*. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016.pdf>.

ACKNOWLEDGMENTS

The authors thank the members of the working group for their help in getting the project started and for providing feedback at various points along the way.

Rachel Bruce, Jisc

Lucie Burgess, University of Oxford

Todd Chavez, University of South Florida

Joy Davidson, University of Glasgow

Tim DiLauro, Johns Hopkins University

Cynthia Hudson-Vitale, Washington University in St. Louis

Leo Konstantelos, University of Melbourne

Sara Mannheimer, Montana State University

Holly Mercer, University of Tennessee

Amy Nurnberger, Columbia University

Sally Rumsey, University of Oxford

Perry Willett, California Digital Library

Fei Yu, University of Queensland

In addition to working group members, the authors would like to thank the following individuals, who helped with describing circumstances in other countries.

Fred Chan, University of Hong Kong

Janet Copsey, University of Auckland

Mary-Jo Romaniuk, University of Manitoba

Titia van der Werf, OCLC Research

INTRODUCTION

Data management underpins current and future research, funder mandates, open access initiatives, researchers' reputations, and institutional ranking. While it is widely recognized that it's necessary to provide data management support, recognition that it requires sustainable funding is slower in coming.

Beagrie, Chruszcz and Lavoie estimate that the costs of data repositories are a magnitude greater than typical institutional repositories that focus on e-publications.¹ These costs primarily occur during acquisition and ingest of datasets and take the form of staffing.²

Even as the community is beginning to understand the costs,³ it must begin to address how data management might be funded. A recent assessment of well-established and well-funded national disciplinary data centers revealed:

- Significant increases in research, teaching and studying efficiency were realized by the users as a result of their use of the data centers;
- The value to users exceeds the investment made in data sharing and curation.⁴

In order to explore the various possibilities, we provide an overview of several funding strategies and their standing in the US. The arguments for and against each strategy are presented and circumstances in other countries are described in the appendix.

Getting Institutional Budget Support

Because research data is a valuable university asset, an institution should build ongoing funding into its base to provide resources to the units responsible for managing that asset. These units may be the library, information technology, other units charged with the tasks, or a combination of units. Data management services are not very different than other services the university directly funds, but they are new, additional services that require new resources. However, as in many other countries, the US economy has suffered, and subsequently higher education budgets have suffered as well.⁵ This makes it less likely that new funding will be made available. It should be recognized that, while data management costs money, not providing these services could result in repeating costly data gathering for subsequent research, could compromise compliance with funders requirements and university policy, could result in questions about the university's research integrity, and ultimately could decrease the federal research dollars the institution receives.

Including Data Curation Costs in Grant Proposal Budgets

For funded research, data management costs should be included in the project budget. Including these costs in the budget ensures that funding agencies that require data management are paying for it. For most funding agencies, costs for data management during the research period (including preparing the data for deposit) can be included in direct costs, but ongoing costs cannot. Even when data management can be included in project budgets, funders vary on the acceptable costs that can be included.

Additionally, it is often difficult for the researcher to “reallocate” grant funds from research activities that they consider more important.

Ongoing costs to manage the data after the project is finished ought to be included in indirect costs. This is the most rational solution, because data management is an indirect cost of research. The problem is that the indirect cost calculation is slow to change, and other parts of the university depend on the money as currently distributed. It is difficult to increase the percentage or to change the distribution so that sufficient funds could go to the units that provide data management services.

It can sometimes help a budget proposal to include data management as a matching amount the university contributes to a project, but that doesn't help the units that provide the services, unless the funds are redistributed to them.

Charging Researchers

Data management units could pass on costs to the researchers or to their departments. Since grant funds typically cannot be used after the grant is over, there are two methods to accomplish this: by charging a set maintenance fee for a particular period, after which ongoing retention is reassessed, or by adopting a pay-once model, which takes into consideration ongoing costs. However, this may not solve the problem, since it merely redistributes the cost to the researchers or their departments.

Charging Data Users

Those who use the data and are likely to benefit from it could reasonably be expected to pay for it. The current mechanism for monetizing datasets typically involves a technology transfer, or technology commercialization, office. However, it is difficult to predict which data will be in most demand. Furthermore, most funder mandates require free public access to the data. This makes it unlikely that charging data users could create a sufficient revenue stream for funding data management services.

Endowment

A philanthropic gift large enough to set up an endowment to cover ongoing costs would solve the problem. However, data management is not a terribly attractive thing to endow. Institutions with a particular center of excellence may be more likely to attract such largess. In addition, some institutions are unable to set up funding that spans fiscal years. Currently, endowments for data management services are rare and specialized. The efficacy of this option remains to be seen.

Repository Funding

Government funding or private funding is sometimes made available to establish digital preservation programs, data curation, or repository development. This money would help to establish a data management operation, but would not help to sustain the service. Many existing data repositories have begun with this type of outside funding, but sustainability requires other funding strategies.

Making Do

Without any, or enough, of the above sources of funding, institutions will have to make do with existing funding. One option is to terminate one or more other services in order to offer data management.

Another option is to “outsource” to external data repositories. Some external repositories charge deposit fees, though these can be included in grant budgets.

While external data repositories can be a good option, many provide access to data but make no effort to meet digital preservation standards. They may assume the data is preserved by the researcher's institution or consider this service as outside their scope. Further, while some disciplines are well-supported by centralized data repositories, other disciplines have few or no options, which again leaves the university to take responsibility for those datasets.

WHO PAYS FOR RESEARCH DATA MANAGEMENT (RDM)?

Funding strategy	Pros	Cons
Institutional budgetary support	<ul style="list-style-type: none"> – Protects data as an institutional asset – Supports all researchers regardless of funding source – Faculty support this option 	<ul style="list-style-type: none"> – Difficulty in garnering administrative support unless additional funding is available – Requires re-allocation of institutional funds
Charging grant budgets: direct funds	<ul style="list-style-type: none"> – Current mechanism in place for most grants for some RDM activities 	<ul style="list-style-type: none"> – Only covers funded researchers – Not allowed to extend past grant expiration – Faculty resist this option
Charging grant budgets: indirect costs or overhead	<ul style="list-style-type: none"> – Most rational choice – Faculty support this option 	<ul style="list-style-type: none"> – Difficulty in garnering administrative support, because the re-negotiation of this rate is very slow and controversial
Charging depositors (faculty or departments)	<ul style="list-style-type: none"> – Costs are borne by those generating the data (and the grant funding) 	<ul style="list-style-type: none"> – May leave out unfunded researchers – Reallocates the problem, rather than resolving it
Charging data users	<ul style="list-style-type: none"> – Costs are borne by those benefitting from the university's data assets – May generate a revenue source from outside the institution 	<ul style="list-style-type: none"> – Only receive payment for high-demand or commercially-relevant data – Most funding agencies strongly encourage or require free public access
Endowment	<ul style="list-style-type: none"> – Relatively stable, ongoing support – Supports all researchers regardless of funding source 	<ul style="list-style-type: none"> – Volume of data could outpace endowment funds – Some institutions are unable to handle payment across fiscal years
Funding for data repository development	<ul style="list-style-type: none"> – Currently available – Helps to get a program started – Supports all researchers regardless of funding source 	<ul style="list-style-type: none"> – Funding is for a limited purpose and a limited-time – Doesn't support ongoing efforts
Making do: finding room in existing budgets	<ul style="list-style-type: none"> – Supports all researchers regardless of funding source 	<ul style="list-style-type: none"> – Must cut other services or programs – Outsourcing is only an option for some disciplines and many don't offer preservation

Funders are increasingly asking for “institutionally-backed solutions.”⁶ They see the permanence of a library preservation solution as preferable to the access nodes that predominate among disciplinary data “repositories.” In one of the most comprehensive surveys of faculty, the University of North Carolina asked, “In your opinion, where should the funding come from to cover the costs of data management and storage for research supported by grants, contracts, or other external sources of funding?” Over half of the 2010 respondents answered that “It should be paid for by the University from overhead/F&A funds it receives from grants and contracts.”⁷

When asked about funding for research not supported by grants, contracts, or other external sources, 63% of respondents thought it should come from university funds.⁸ This indicates that institutional funds, whether they derive from the overhead from external grants or other revenue streams, is a preferred option for many faculty.

In our own informal survey of institutional data repositories,⁹ eighteen of the twenty respondents reported that the library’s base budget covers at least some of the expenses, with seven saying that is their only source of funding. Seven fees from researchers and four fees from departments. Five get institutional funding specifically for data management. Four get money from the IT budget. Only one institution reported getting direct funds from grant-funded projects and only one reported getting indirect funds from grant-funded projects. None reported getting fees from users, having an endowment, or having had a grant to fund development of the repository. In general, most institutional data repositories rely on their library budgets, although a few are experimenting with multiple funding sources.

It is very possible that a combination of these approaches will be necessary. Initially most institutions begin with one or two sources of funding and as services grow, they explore other options. As data management service providers, we should continually be aware of our funding options and engage our institution’s leadership in the conversation¹⁰ regarding the ideal funding approach for long-term sustainability. To determine their best course of action, each institution will need to balance the costs of providing data management services with the benefits of responsible stewardship of their institutional assets.

APPENDIX: CIRCUMSTANCES IN OTHER COUNTRIES

AUSTRALIA

- There are no funder mandates for data sharing or data management plans as yet, though they are implicitly encouraged by Research Councils.
- There is no funding body-endorsed or government-endorsed service for storing and managing data sets, but the government provides \$50m through the Research Data Storage Infrastructure project,¹¹ which many institutions use to make data accessible.
- A national data registry of metadata descriptions (Research Data Australia)¹² was funded by the Australian Government via the Australian National Data Service.¹³

CANADA

- Canada's federal funders are currently proposing a mandate for data management plans in grant proposals.¹⁴ Two of the funders have policies about depositing and archiving data sets.
- There is no government/national archives to support data curation.
- CARL (the Canadian Association of Research Libraries) is working with regional consortia to build a national network of experts and digital preservation infrastructure that will be available to all Canadian institutions.¹⁵

HONG KONG

- At present, research funders have no mandate for data management plans or data sharing.
- The Research Grants Council,¹⁶ the major research funder, indicates that additional weight will be given when applicants are willing to share research data.
- There are currently no government or institutional data archives.

NETHERLANDS

- NWO (Netherlands Organization for Scientific Research) has a pilot of seven rounds of funding¹⁷ that requires applicants to explain how they will make their research data findable and suitable for re-use; successful applicants must submit a data management plan.
- Most universities provide data management support to their researchers, which is often provided by library staff.
- Long-term data management costs are subsidized by the government through the financing of three national data archiving centers: 3TU.Datacentrum,¹⁸ DANS,¹⁹ and SURFSara.²⁰

NEW ZEALAND

- There are no funder mandates for data management plans as yet.
- There is no government service for storing and managing data sets from research institutions although there is a metadata harvesting service for research papers and related resources.
- Discussions are underway at the national level in regard to the need to consider a national preservation service for a range of data, including research data.

UK

- Jisc²¹ is brokering national shared services to support more effective and efficient institutional investment in managing and sharing research data.
- Jisc is actively involved in a number of European Commission projects such as FOSTER,²² OpenAIRE2020²³ and EUDAT2020²⁴ to develop international RDM support and infrastructure.²⁵
- Research Councils UK²⁶ maintain that the use of public funds to support efficient and cost-effective data sharing is an acceptable cost to include in grant proposals, either as in-project expenses or as indirect costs.
- While it is unlikely that UK Higher Education Institutions would charge academic researchers to access and reuse data generated through public funding, there may be possibilities for charging for commercial reuse

NOTES

1. Beagrie, Neil, Julia Chruszcz and Brian Lavoie. 2008. *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities: Final Report—April 2008*. Bristol, England: HEFCE.
<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>.
2. Beagrie, Neil, Brian Lavoie and Matthew Woolard. 2010. *Keeping Research Data Safe 2: Final Report—April 2010*. Bristol England: HEFCE.
<http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>.
3. 4C: Collaboration to Clarify the Costs of Curation. 2015. "Investing in Curation: A Shared Path to Sustainability - the 4C Roadmap." Accessed 24 July 2015. <http://4cproject.eu/roadmap>.
4. Beagrie, Neil, and John Houghton. 2014. *The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres*. Bristol, England: Charles Beagrie Ltd, Victoria University, and Jisc.
5. Mitchell, Michael, Vincent Palacios and Michael Leachman. 2014. *States are Still Funding Higher Education Below Pre-recession Levels*. Washington, DC: Center on Budget and Policy Priorities.
<http://www.cbpp.org/sites/default/files/atoms/files/5-1-14sfp.pdf>.
6. Rinehart, Amanda. 2015. "Data Tales from the Hills: The Request for Information (RFI)." *Databrarians* (blog). Posted 24 June. <http://databrarians.org/2015/06/data-theses-from-the-hills-the-request-for-information-rfi/>.
7. Provost's Task Force on the Stewardship of Digital Research Data. 2012. *Research Data Stewardship at UNC: Recommendations for Scholarly Practice and Leadership*. Chapel Hill, NC: University of North Carolina.
http://sil.unc.edu/sites/default/files/general/research/UNC_Research_Data_Stewardship_Report.pdf.
8. Ibid.
9. Erway, Ricky. 2015. "What is Actually Happening out There in Terms of Institutional Data Repositories?" *hangingtogether.org* (blog). Posted 27 July 2015. <http://hangingtogether.org/?p=5342>.
10. Erway, Ricky. 2013. *Starting the Conversation: University-wide Research Data Management Policy*. Dublin, Ohio: OCLC Research.
<http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf>.
11. CQ University Australia. "Research Data Storage Infrastructure (RDSI) Project." In: eResearch/Data Services. Accessed 11 November. <https://my.cqu.edu.au/web/eresearch/research-data-storage-infrastructure>.
12. <https://www.rds.edu.au/about>.
13. <https://researchdata.ands.org.au/>.
14. Government of Canada. 2015. "Draft Tri-Agency Statement of Principles on Digital Data Management." In: Science.gc.ca. Last modified 20 July 2015.
<http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>
15. <http://www.carl-abrc.ca/en/scholarly-communications/research-collections.html>.
16. <http://www.ugc.edu.hk/eng/rgc/index.htm>.
17. NWO (Netherlands Organization for Scientific Research). "Start Pilot Data Management." In: Policies/Data Management. Last updated 7 May 2015.
<http://www.nwo.nl/en/policies/open+science/data+management>.

18. <http://datacentrum.3tu.nl/en/home/>.
19. <http://www.dans.knaw.nl/en>.
20. <https://www.surf.nl/en/about-surf/subsidiaries/surfsara/>.
21. <https://www.jisc.ac.uk/>.
22. <http://www.rcuk.ac.uk/>.
23. OpenAIRE. "OpenAIRE2020: A New Horizon for Open Science." News & Events. 02 December 2014. <https://www.openaire.eu/openaire2020-press-release>.
24. http://cordis.europa.eu/project/rcn/194928_en.html.
25. Jisc. "Jisc Invests in Collaborative Projects to Improve Research Data Management." Jisc News. 24 March 2015. <https://www.jisc.ac.uk/news/jisc-invests-in-collaborative-projects-to-improve-research-data-management-24-mar-2015>.
26. <http://www.rcuk.ac.uk/funding/>.

For more information about our work related to data management, please visit:
www.oclc.org/research/themes/research-collections/datacuration.html



6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878

T: +1-614-764-6000

F: +1-614-764-6096

www.oclc.org/research

ISBN: 1-55653-501-5
978-1-55653-501-7
1512/215544, OCLC